# Hybrid ARIMA-ANN Model for Solving Nonlinearity In Time Series Data

Widyanti Rahayu[1,]* Vera Maya Santi[1], Dania Siregar[1], Dimas Kuncoro Djati[1]

[1] Jakarta State University, Jakarta, Indonesia
wrahayu@unj.ac.id

**Abstract.** Shallots (*Allium cepa L. var. aggregatum*) are one of the most important culinary spices in the world. Shallots contain vitamin C, potassium, dietary fiber, folic acid, calcium and iron. Because the nutritional content and benefits of shallots are very complete, the demand for shallots is quite high in various regions in Indonesia. It takes an evaluation process of shallot price data to determine the future movement of shallot prices through the right forecasting method to obtain accurate shallot price predictions. In this study, the forecasting was conducted using the Hybrid Autoregressive Integrated Moving Average - Artificial Neural Network (ARIMA-ANN) method, because the data has a non-linear pattern. The results of this forecasting can be used as a reference to predict shallot prices for the next few periods. The analysis steps are data preparation, modeling and forecasting using ARIMA, forecasting residual of ARIMA model using ANN method, and then, forecasting results are obtained from the difference between actual data values and residual forecasting results.The comparison results with ARIMA method show that forecasting using Hybrid ARIMA-ANN method produces a good accuracy value. Data models with ARIMA and Hybrid ARIMA-ANN to predict shallot prices in this study have MAPE 19.81% and 19.1%, respectively. Based on these results, the Hybrid Autoregressive Integrated Moving Average - Artificial Neural Network (ARIMA-ANN) method is very well used to forecast shallot prices.

**Keywords:** ARIMA, shallot prices, Hybrid ARIMA-ANN, MAPE

## 1. Introduction

Food commodities play an important role as a contributor to inflation in a region. In order to achieve stable commodity prices in Indonesia, the government has regulated commodity prices in Law no.7 of 2014. Shallots (*Allium cepa L. var. aggregatum*) are one of the important commodities in Indonesia and are the main cooking spices in the world. Because of its complete benefits, the demand for shallots is quite high in Indonesia. Studying shallot price data using appropriate forecasting methods is very useful in obtaining accurate shallot price predictions. In general, analysis for time series data can use the autoregressive moving average (ARIMA) model. The ARIMA model is relatively good at modeling linear and stationary time series data. This is as shown in Saumi's research [7] regarding forecasting the number of claims for the JHT program


N. Djam'an et al. (eds.), *Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023)*, Advances in Computer Science Research 109,
https://doi.org/10.2991/978-94-6463-332-0_9

at BPJS employment in Langsa city. Some studies to predict shallot prices data using the ARIMA model, by Windhy, et al. [8] and Afridar, et al.[1].

Based on BPS data in the consumption sector, especially commodity price data, has an irregular shape due to rapid fluctuations over time. The ARIMA model cannot provide an appropriate model if the data do not satisfy the linear assumption. If the non-linear timeseries data is still modeled using ARIMA, it will produce quite large errors, which makes the model less suitable for use in forecasting. Zhang [10] said that the ARIMA model requires a model that can capture nonlinear patterns. One method for forecasting non-linear time series data is to use the artificial neural network (ANN) method. The ANN method is a method that creates a model by imitating the human nervous system, Pitts [6] . The ANN method processes information used for the analysis of time series data containing nonlinear components of ARIMA. Various studies have been conducted to overcome the problem of nonlinearity in time series data. Khalil, et al.[5] compared the ANN model with the ARIMA model in Turkey's monthly aluminium exports to Iraq, resulting in a better ANN model than the ARIMA model.

The characteristics of shallot price data are thought to contain linear and nonlinear patterns, so forecasting using a combination of the ARIMA and ANN models is expected to produce better forecasts than the ARIMA model. In this research, a study will be carried out on the characteristics and performance of the hybrid ARIMA-ANN model for forecasting shallot prices in Indonesia, which is expected to produce an accurate statistical model.

## 2. METHODS

### 2.1. Materials.

The data used in this study are daily shallot prices from October,1 2022 to September,30 2023. These data are secondary data from Info Pangan Jakarta which are officially announced.

### 2.2. Work Procedure

**Preprocessing Data.** Data preprocessing is the first step in processing data, to obtain data that is ready to be processed. First check stationarity of data. Stationarity tests were performed using the Augmented Dickey Fuller (ADF) test and Levene test. Two types of non-stationary in the data, non-stationary on the variance and non-stationary on the mean. Suitable transformation and differencing can stabilize the variance and make the data stationary in the mean.

**Build an ARIMA model.** The process will continue with the identification of the ARIMA model to determine the best ARIMA order. ARIMA first introduced by statisticians George Box and Gwilym Jenkins in 1970, they combine Autoregressive (AR) and Moving Average (MA), and adding new Integrated component to make data stationary. This process is carried out by looking at the ACF and PACF plots to determine the ARIMA order. Furthermore, the parameter estimation process and diagnostic tests from the predetermined model are carried out. The next process is added step function intervention from the selected ARIMA model and hypotheses were tested for the parameters.

**Build an ANN model**. Artificial neural networks are one of such models that are able to approximate various non linearities in the data. ANNs are Dexible computing frameworks for modeling a broad range of nonlinear problems. One significant advantage of the ANN models over other classes of nonlinear model is that ANNs are universal approximators which can approximate a large class of functions with a high degree of accuracy. Their power comes from the parallel processing of the information from the data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data.

**Build the Hybrid ARIMA-ANN.** Both ARIMA and ANN models have achieved successes in their own linear or nonlinear domains. However, none of them is a universal model that is suitable for all circumstances. The approximation of ARIMA models to complex nonlinear problems may not be adequate. On the other hand, using ANNs to model linear problems have yielded mixed results. These data are secondary data from Info Pangan Jakarta which are officially announced.

## 3. RESULTS AND DISCUSSION

The first step in this procedure is to check stationary data. Based on the data processing, we obtained results shown in Table 3.1.

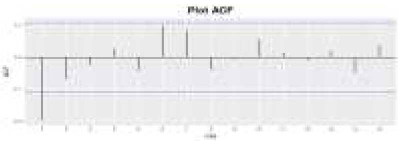
Table 3.1. Stationarity Test results

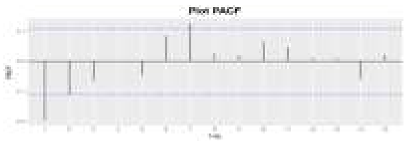
| Test | Prob I | Prob II |
|---|---|---|
| ADF test | 0.809 | 0.012 |
| Levene test | 2.2e-16 | 0.5544 |

From the result in Table 3.1., it can be seen that the Levene test result get p-value 2.2e-16, so the data is not stationary in variance. Next, we check stationary in the mean with the ADF test and get p-value = 0.809, so the data is not stationary in the mean. To make the data stationary, we use Box-Cox transformation with lambda 0.5 and use once differencing. Next we check stationary again, it can be seen that the ADF test result get p-value = 0.012, that means it's stationary in mean. Next, using Levene test and it get p-value = 0.5544, that means it's stationary in variance. To train the ARIMA model, the data must be divided into training data and test data. For training data we used 90% of the data and 10% was used as test data.


Fig. 3.1. Time Series Plot Data Train and Test

From data training we obtained ARIMA order from ACF and PACF plot.




From the ACF and PACF plot, we get MA(1) and AR(2), also there is a seasonality in 7 days period. Next, we compare 5 ARIMA model include overfitting and seasonality, as seen below.

Table 3.2. Comparison of 5 ARIMA Models

| ARIMA | ME | RMSE | MAE | MPE | MAPE |
|---|---|---|---|---|---|
| (3,1,3)(1,1,0) | -2460.406744 | 2888.234305 | 2466.314393 | -17.47773374 | 17.510554 |
| (2,1,3)(0,1,1) | -2535.55557 | 2914.432355 | 2543.213441 | -17.96705527 | 18.009599 |
| (3,1,0)(0,1,1) | -2604.81451 | 2971.357982 | 2605.248277 | -18.42813442 | 18.430544 |
| (3,1,1)(0,1,1) | -2607.380198 | 2973.532348 | 2607.687583 | -18.44523338 | 18.446941 |
| (3,1,3)(0,1,1) | -2627.918665 | 2991.319971 | 2627.918665 | -18.58246147 | 18.582461 |

From Table 3.2., the results of the best ARIMA model for shallot data is ARIMA (3,1,3)(0,1,1). To train the new ANN model we used the residual from ARIMA (3,1,3)(0,1,1).

From ARIMA model and ANN model we obtain hybrid ARIMA - ANN model. An hybrid ARIMA-ANN model for 5 models as seen below.

Table 3.3. Comparison of 5 Hybrid ARIMA-ANN Models

| Orde | ME | RMSE | MAE | MPE | MAPE |
|---|---|---|---|---|---|
| (3,1,3)(1,1,0) 9-1-1 | -2170.534 | 2601.778 | 2213.872 | -15.4652 | 15.7112 |
| (3,1,3)(1,1,0) 2-9-1 | -2234.637 | 2690.656 | 2255.580 | -15.952 | 16.0708 |
| (3,1,3)(1,1,0) 2-8-1 | -2275.306 | 2723.462 | 2291.261 | -16.221 | 16.311 |
| (3,1,3)(1,1,0) 2-7-1 | -2318.431 | 2760.988 | 2331.411 | -16.515 | 16.587 |
| (3,1,3)(1,1,0) 3-5-1 | -2347.173 | 2791.987 | 2356.222 | -16.717 | 16.768 |

From the table 3.3. above it can be seen that the best model for the hybrid ARIMA-ANN is (3,1,3)(1,1,0) 9-1-1. The prediction results between ARIMA and hybrid ARIMA-ANN are shown in the following graph.

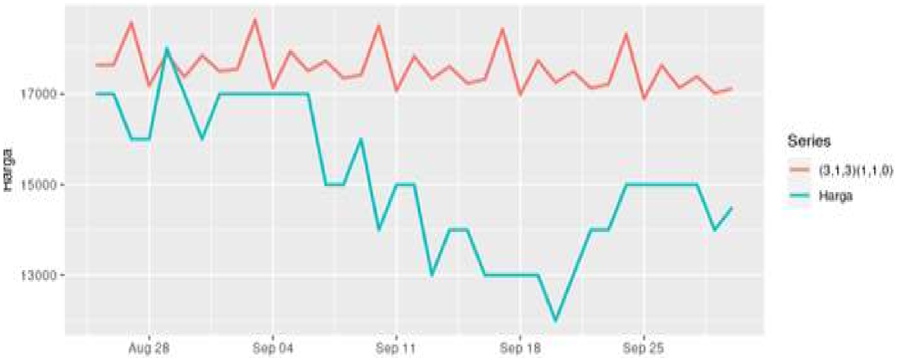


Fig. 3.2. ARIMA Prediction Plot

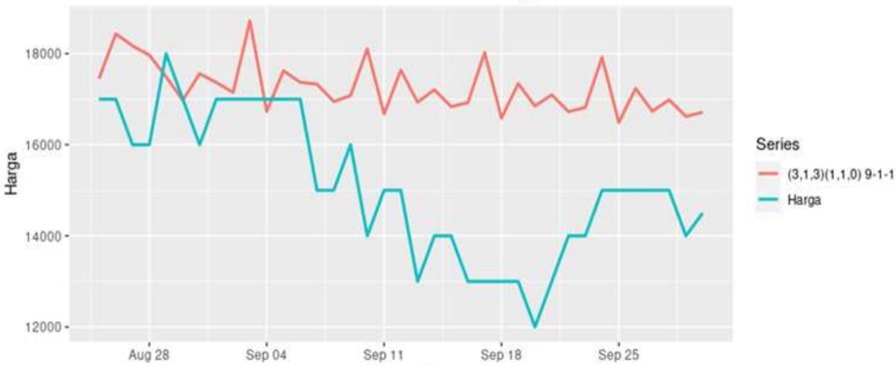


Fig. 3.3. Hybryd ARIMA-ANN Prediction Plot

From the two graphs above, it can be seen that the ARIMA-ANN model graph is closer to the actual data values than the ARIMA model. This can be seen from the comparison of MSE and MAPE values in the following table.

Table 3.4. Comparison of ARIMA and Hybrid ARIMA-ANN

| | ME | RMSE | MAE | MPE | MAPE |
|---|---|---|---|---|---|
| ARIMA | -2460.4067 | 2888.2343 | 2466.3143 | -17.47773 | 17.51055 |
| Hybrid ARIMA-ANN | -2170.5345 | 2601.7782 | 2213.8724 | -15.46529 | 15.71125 |

From table 3.4. above, it can be seen that the hybrid ARIMA-ANN model has smaller RMSE and MAPE values than ARIMA model. So the ARIMA-ANN hybrid model is a more suitable model for predicting shallot price.

# 4. CONCLUSION

Shallots (*Allium cepa L. var. aggregatum*) are one of the most important culinary spices in the world. To predict the future prices we need a good forecasting model. Most used model for forecasting is ARIMA, but ARIMA has weakness that it can’t be used in nonlinear problem, so we used Hybrid ARIMA-ANN proposed by Zhang et. all (2003). The model combine the output of ARIMA and using the residual of the ARIMA model to train ANN model to capture non linear pattern and then combine it together. From our testing we get that hybrid ARIMA-ANN model is better with RMSE 2601 and MAPE 15.71%, than ARIMA model that have RMSE 2888 and MAPE 17.51%.

# References

1. Afridar, H., Gunawan, Andriani, W.: Penerapan Metode ARIMA untuk Prediksi Komoditi Harga Bawang Merah di Kota Tegal. Indonesian Journal of Informatics and Research, 3(2), 18-29. (2022)
2. Azevedo, F.A., Carvalho, L.R., Grinberg, L.T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Jacob Filho, W., Lent, R., & Herculano-Houzel, S.: Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. The Journal of Comparative Neurology, 513(5), 532–541. (2009).
3. Estiko, F. I., & Wahyuddin, W. : Analysis of Indonesia’s Inflation Using ARIMA and Artificial Neural Network. Economics Development Analysis Journal, 8(2), 151–162. (2019).
4. Fausset, L.: Fundamental of Neural Network (Archetectures, Algorithms, and Applications). Upper Saddle River, New-Jersey: Prentice-Hall. (1994).
5. Khalil, D.M., and Hamad, S. R.: A Comparison of Artificial Neural Network Models and Time Series Models for Forecasting Turkey's Monthly Aluminium Exports to Iraq, Journal of Survey in Fisheries Sciences, 10(4S), 1536-1557. (2023)
6. McCulloch, W. S., & Pitts, W. : A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 99-115.(1990).
7. Saumi, F.: Penerapan Model Arima Untuk Peramalan Jumlah Klaim Program Jaminan Hari Tua pada BPJS Ketenagakerjaan Kota Langsa, Barekeng, 14(4), 491-500. (2020)

8. Windhy, A.M., Suci, Y.T. ., & Jamil, A.S.: Analisis Peramalan Harga Bawang Merah Nasional Dengan Pendekatan Model Arima. Seminar Nasional Pembangunan Pertanian Berkelanjutan Berbasis Sumber Daya Lokal, 591-604 (2019).
9. Susila, M.R., Jamil, M., & Santoso, B.H.: Akurasi Model Hybrid ARIMA-Artificial Neural Network dengan Model Non Hybrid pada Peramalan Peredaran Uang Elektronik di Indonesia. Jambura Journal of Mathematics, 5(1), 46–58. (2023)
10. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. Elsevier 50, 159-175 (2003)
11. Info Pangan Jakarta Homepage, https://infopangan.jakarta.go.id/publik/report_commodity last accessed 2023/10/2.